**Abstract Title Page**

**Title: Compared to what? Estimating Causal Effects for Latent Subgroups to Understand Variation in the Impacts of Head Start by Alternate Child Care Setting**

**Authors and Affiliations:**

Avi Feller
Harvard University
Department of Statistics
avi.feller@gmail.com

Todd Grindal
Abt Associates
toddgrindal@gmail.com

Luke Miratrix
Harvard University
Department of Statistics
lmiratrix@fas.harvard.edu

Lindsay C. Page
University of Pittsburgh
School of Education
lpage@pitt.edu

(Authors in alphabetical order)

## Abstract Body

**Background / Context:**
Head Start programs currently provide early childhood education and family support services to more than 900,000 low-income children and their families across the United States with an annual budget of around $8 billion in state and federal funds. Researchers and policy makers have debated the program's effectiveness since its inception in 1964. Dozens of studies (see for example Currie & Thomas, 1993; Deming, 2009; Ludwig & Miller, 2007) over the last four decades have provided a mixed picture of results with some showing strong positive impacts and others indicating small or null effects. In an effort to resolve these long standing disputes over the impact of Head Start, the Administration for Children and Families commissioned a random assignment study of Head Start programs in 2000. To date, the results of the Head Start Impact Study (HSIS) have done little to settle debates about the effects of Head Start attendance on children's development.

Some have suggested that the HSIS results may be attributable, in part, to the fact that more than one-third of children assigned to the study control group enrolled in some form of early childhood education that offered services similar to those provided by Head Start (National Forum on Early Childhood Policy and Programs, 2010). There is some empirical support for this hypothesis. A recent meta-analysis of 28 studies of Head Start conducted between the program's inception and 2007 found that much of the variation in the findings regarding Head Start's impact on child achievement and cognitive development could be explained by differences in the types of preschool services used by the control group (Shager et al, 2013).

However, unpacking treatment effect by variation in control group setting is difficult using traditional methods. The standard Two-Stage Least Squares (TSLS) framework can be used to estimate effects of Head Start relative to no Head Start among students who are "compliers" with the randomization scheme. Yet, there are many other instances of uncontrolled variation that are of policy interest in the context of HSIS. Specifically, in HSIS, both treatment and control group students ended up in different types of care settings such as Head Start, other center care, parent care, and home daycares. Furthermore, many students crossed over their assignment, ending up in Head Start centers when assigned to control, and not enrolling in Head Start even when offered the opportunity (Table 1). Given this variation, an important question to ask is whether enrollment in Head Start by virtue of the randomized offer to do so has differential effects according to the alternate care type setting in which a child would otherwise be. Answers to this question are policy relevant, as they may inform future efforts to target high quality, comprehensive services such as Head Start to those children and families who potentially stand to benefit most from participating.

**Purpose / Objective / Research Question / Focus of Study:**
In this paper, we extend the main HSIS findings to examine whether the offer of Head Start was differentially beneficial for students grouped by the care setting they would have received absent the offer of Head Start. In order to do so, we develop a novel methodological approach that connects principal stratification as an analytic framework with a hierarchical random effects model and Bayesian estimation strategies to stratify children into latent subgroups defined by the pair of child care settings each child would experience under the randomized offer of Head Start

and under no such offer. That is, we assign students to a subgroup defined not only by the care type we observe them to have taken up but also by the care type they would have taken up under the counterfactual experimental condition. By examining treatment-control differences in outcomes only for those children who, for example, would enroll in Head Start if offered the opportunity, but who otherwise would be cared for by a parent, we are able to able to estimate the causal impact of Head Start across the alternate care-type settings. Our analyses will focus on outcomes such as such as performance on the Peabody Picture Vocabulary Test and the Woodcock-Johnson III assessments gathered for several years after the initial HSIS randomization.

**Setting:**
We use secondary data from the HSIS, the first randomized trial of Head Start in the program's history.

**Population / Participants / Subjects:**
In total, 4,440 children in 202 center groups were randomized to treatment or control in the original HSIS. Children in our sample were diverse in their background characteristics – 30% were Black, 38% were Hispanic, 30% spoke a non-English home language, 50% lived with both biological parents, 19% had a mother who was a recent immigrant, and 50% were male.

**Intervention / Program / Practice:**
Children randomized to treatment were offered a seat in a classroom in a Head Start program in fall 2002 for the 2002-2003 school year. In total, 73% took up the offered slot and enrolled in Head Start in the treatment year. Approximately 7% of children assigned to treatment enrolled in a non-Head Start center, 9% experienced parent care, were cared for by a relative, or enrolled in family daycare (around 11% have an unknown care type). Children assigned to treatment who took up the offered Head Start treatment were enrolled in programs that emphasized a "whole child" approach, meaning the program targeted children's cognitive, academic, and socio-emotional skills, as well as the health and nutrition of enrolled children grade (U.S. Department of Health and Human Services, 2010). Head Start programs typically emphasized parent involvement and offered a wide array of comprehensive services for families.

Children randomized to control conditions were free to take up any available early childhood program except for that provided by the Head Start to which they had applied and had not won a seat. In practice, 11% of control group children enrolled in Head Start centers (most in the very centers in which they had lost a lottery), 25% enrolled in a non-Head Start center, and 44% experienced parent care, were cared for by a relative, or enrolled in family daycare (around 21% have an unknown care type).

**Significance / Novelty of study:**
The HSIS was a very large-scale intervention that has received much attention. The novelty of this particular study is in incorporating both a random effects model and a principal stratification framework into a single analysis. Furthermore, our principal stratification framework includes multiple strata for the "compliers," something not often implemented, allowing insight into the difference in treatment impacts among subgroups. To the best of our knowledge, we are the first

to implement a principal stratification analysis with both random effects for clusters and multiple subgroups for compliers.

**Statistical, Measurement, or Econometric Model:**
We approach this problem by combining a principal stratification framework (Frangakis and Rubin, 2002), which is common in the medical trials literature, with a random effects hierarchical model, which is common in the education literature. In particular, we model the population of students as a mixture of latent subgroups, with each subgroup defined by the pair of what care type a child would receive with and without an offer of Head Start. For example, one latent subgroup is "Center Compliers," children who would attend Head Start if offered, and would otherwise receive care in a non-Head Start center. Another is "Always Head Start," those who would attend a Head Start center regardless of an offer or not. Group membership is not fully observed because it depends on the counterfactual action that a student would take under the alternative randomly assigned condition. Note that the classic Instrumental Variables (IV) model is simply a special case of the richer principal stratification framework, where the latent subgroups are the more familiar Always Takers, Never Takers, and Compliers.

Our framework consists of two key sub-models. The first is the *choice sub-model*, which predicts each child's care choice under both treatment and control based on a robust set of baseline covariates. This has the important additional benefit of providing information on relevant predictors of child-care choice. The second is the *outcome sub-model*, which is effectively a classic Intent-to-Treat (ITT) analysis for each of the principal strata predicted in the first sub-model. Both sub-models allow for random effects for each of the 202 center groups in the study. We jointly fit these two sub-models via Hamiltonian Monte Carlo (HMC) using the open-source Bayesian modeling platform, Stan (Hoffman & Gelman, 2013).

In short, we stratify children into latent subgroups defined by the pair of what childcare setting each child would experience under the randomized offer of Head Start and under no such offer. By then examining treatment-control differences in outcomes for only those children who, for example, would enroll in Head Start if offered the opportunity, but who otherwise would be cared for by a parent, we can estimate the impact of Head Start across the alternate care-type settings. In particular, we compare performance on outcomes such as the Peabody Picture Vocabulary Test (PPVT) and Woodcock-Johnson III assessments for a series of years following the assignment of treatment.

Our analytic approach has several main advantages. First, we jointly obtain valid estimates of the causal effects of interest within each principal stratum, allowing for detailed evaluations of effects without the typical multiple comparison issues common in Frequentist models (Gelman et al, 2012). Second, we can incorporate substantive, domain-knowledge-based assumptions directly into the model, such as encoding restrictions on the plausible principal strata. Third, we can directly address the differential attrition of students from the treatment and control groups in the HSIS, as well as a range of other missing data issues, by jointly imputing the missing data as part of the model. Finally, by including random effects for the center groups of random assignment, we can correctly account for the complex HSIS experimental design.

**Usefulness / Applicability of Method:**

We demonstrate the effectiveness of this method by first comparing the results of a simpler Instrumental Variable model to moment-based Instrumental Variable techniques. These comparisons demonstrate that the method does appear to recover overall estimates that are in alignment with other techniques. We then demonstrate the applicability of the method by analyzing the Head Start data to estimate treatment impact on latent subgroups of direct interest to public policy. Estimating these impacts is not otherwise tractable using conventional methods. We finally investigate coverage properties and sensitivity to misspecification by simulation study given a known data generating process for the outcomes and latent subgroups built from a prior estimation step on the raw data. This allows us to examine performance in near-real-world contexts.

**Research Design & Data Collection and Analysis:**
We utilize data from the Head Start Impact Study.

**Findings / Results:**
Our preliminary results show that the effect of Head Start indeed differs by alternative care type. First, as shown in Figures 1 and 2, we find that Head Start has a larger effect on key cognitive outcomes relative to non-center care than to center care. In particular, for almost all models explored, the effect for Non-Center Compliers stochastically dominates effects for Center Compliers; the only non-zero effects we find are for the Non-Center Complier group. Second, there is moderate evidence the effect of Head Start on PPVT appears to persist for Non-Center Compliers, though these effects are small in later years (Figure 1). By contrast, we find little evidence of persistence for the WJ-III assessments (Figure 2). Finally, we find suggestive evidence that there is no beneficial effect of Head Start relative to other Center care.

**Conclusions:**
Our preliminary research suggests that alternative care type is an important source of variation in the Head Start Impact Study, and that the effects of Head Start are primarily for those children who would not otherwise receive formal childcare. We reach this conclusion by utilizing a powerful and flexible modeling approach that can applied to a broad range of interventions, especially those settings in education that could benefit from hierarchical modeling.

Importantly, however, as with all novel modeling approaches, these initial results should be interpreted with caution. In general, generating the results presented here required careful modeling and extensive sensitivity checks. At SREE, we will present a more systematic approach to these checks and ensure that these results indeed hold up to scrutiny. Moreover, we will discuss the results of ongoing technical work that will provide a more complete understanding of the connection between random effects and the discrete choice model. Finally, we hope to share extensive example code for both Instrumental Variables and principal stratification models using Stan, so that researchers can implement our framework on studies with a similar structure.

In the end, we believe that our approach has been useful for highlighting important variation in the effects of Head Start and will prove useful in a range of other settings as well, especially those with a hierarchical data structure.

# Appendices

## Appendix A. References


Bitler, M. P., Domina, T., & Hoynes, H. W. (2012). Experimental evidence on distributional effects of Head Start. Working paper. Available from: http://www.econ.ucdavis.edu/faculty/hoynes/working_papers/bdh-hsis-paper.pdf.

Currie, J., & Thomas, D. (1993). *Does Head Start make a difference?* National Bureau of Economic Research Working Paper, No. 4406. Cambridge, MA.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 111-134.

Frangakis, C. E. & Rubin, D. B. (2002). 'Principal Stratification in Causal Inference. *Biometrics*, (58), 1–29.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.

Hoffman, M. D. & Gelman, A. (2013). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Working paper. Available from: http://www.stat.columbia.edu/~gelman/research/published/nuts.pdf.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*(1), 159-208.

National Forum on Early Childhood Policy and Programs. (2010). Understanding the Head Start impact study. Available from http://www.developingchild.harvard.edu/

Shager, H. M., Schindler, H. S., Magnuson, K. A., Duncan, G. J., Yoshikawa, H., & Hart, C. M. (2013). Can Research Design Explain Variation in Head Start Research Results? A Meta-Analysis of Cognitive and Achievement Outcomes. *Educational Evaluation and Policy Analysis*, *35*(1), 76-95.

## Appendix B. Tables and Figures

### Table 1. Percent of HSIS Children by Care Type

| | 3 Year-Olds | | 4 Year-Olds | |
|---|---|---|---|---|
| | Control | Treated | Control | Treated |
| **Head Start** | 14 | 84 | 11 | 78 |
| **Center Care** | 21 | 5 | 29 | 9 |
| **Other Care** | 15 | 1 | 10 | 1 |
| **Parent Care** | 35 | 5 | 33 | 4 |
| **Unknown** | 15 | 5 | 17 | 7 |

**Figure 1. Impact of the offer of Head Start on PPVT scores for all compliers, non-center compliers, and center-compliers, by age at time of randomization.**

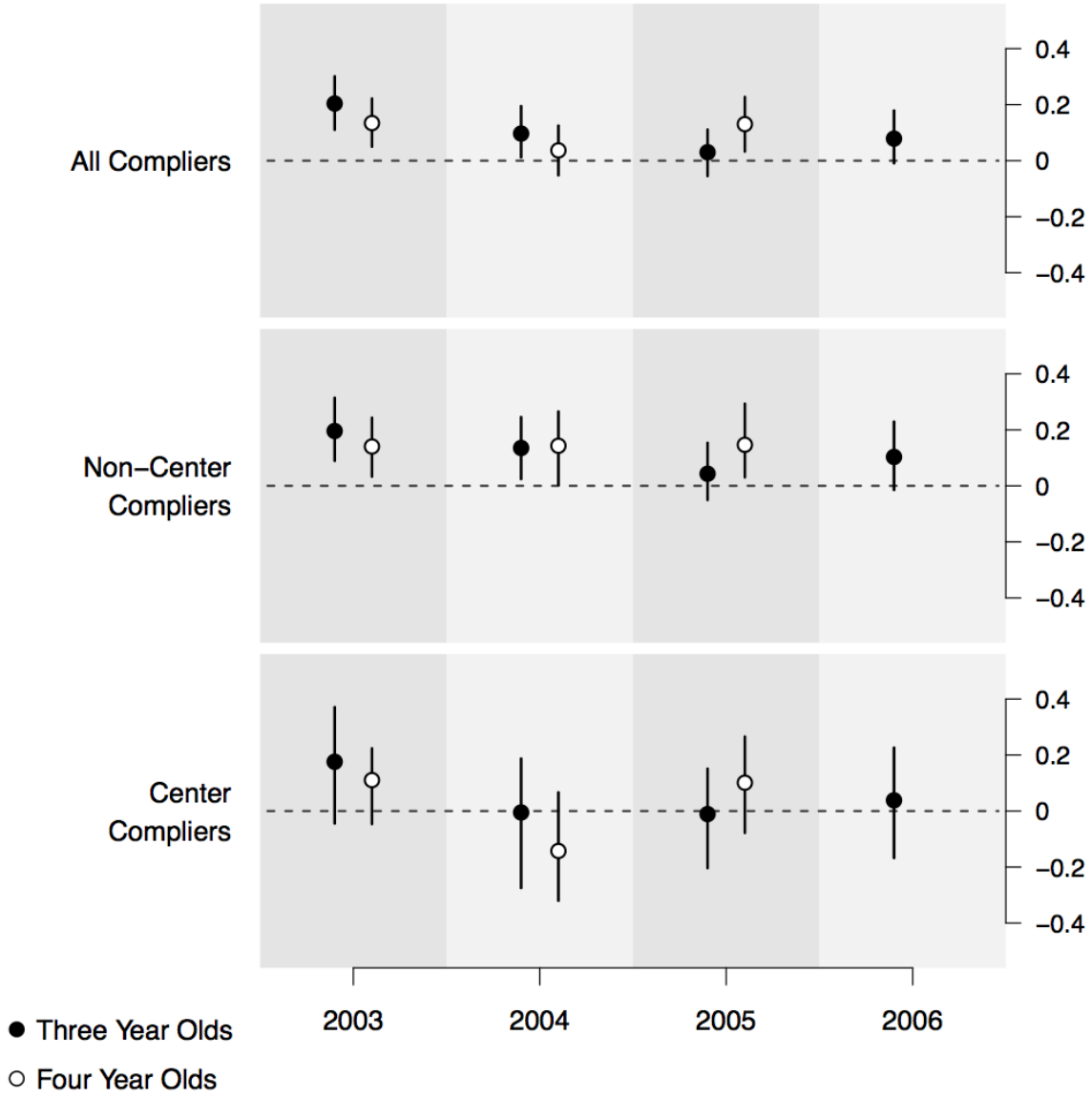

**Figure 2. Impact of the offer of Head Start on WJ-III scores for all compliers, non-center compliers, and center-compliers, by age at time of randomization**